

# UN ESTUDIO COMPARATIVO DE ALGUNOS ESTIMADORES DEL ÍNDICE DE VALOR EXTREMO\*\*

## RESUMEN

Presentamos un análisis comparativo de manera teórica y práctica de algunos métodos elegidos para estimar el índice de valor extremo (IVE) para distribuciones tipo-Pareto. Como era de esperarse, los resultados de la simulación muestran que no existe un método óptimo. Sin embargo la técnica extended GPD (EGPD) ofrece mejoras comparado al estimador de Hill. También aplicamos los métodos elegidos al caso “Danish Fire data” para obtener su índice de valor extremo y un estimador de cuantil alto.

**Palabras clave:** Teoría de Valor Extremo (EVT), Índice de Valor Extremo (IVE), estimador de Hill, Extended GPD (EGPD).

## ABSTRACT

We present a theoretical and practical comparative analysis of various selected methods to estimate the extreme value index for Pareto-type distributions. As it was expected, the simulation results show that there is no best method. However, extended GPD (EGPD) method offers improvements compared to the Hill estimator. We also apply the selected methods to the Danish Fire data in order to obtain its extreme value index and a high quantile estimator.

**Keywords:** Extreme Value Theory (EVT), Extreme Value Index (EVI), Hill estimator, Extended GPD (EGPD).

JEL Classification: C15, C16, G32

\*\* Andrés Mora, profesor – investigador del Colegio de Estudios Superiores de Administración (CESA), Correo electrónico: amora@cesa.edu.co.

## Introducción

La regulación moderna de administración de riesgos requiere estimados de pérdidas potenciales con altos niveles de confiabilidad. Así que es necesario modelos matemáticos sofisticados y herramientas computacionales. Es entonces que surge la teoría del valor extremo (EVT, por sus siglas en inglés) como una disciplina estadística, el cual desarrolla técnicas y modelos para describir los resultados inesperados, anormales o extremos. EVT ha sido aplicado en áreas de ingeniería y recientemente se ha convertido en herramienta fundamental en aplicaciones para finanzas y seguros. Esta teoría brinda modelos para extrapolar datos observados a niveles no observados, debido a que los valores extremos son escasos. De esta manera uno puede decir que EVT se enfoca en el análisis de las colas de la distribución de pérdidas para medir grandes pérdidas que no son tan frecuentes.

Existen dos clases de modelos para tratar valores extremos: *block maxima* (máximos por bloque) y *POT* (picos sobre el umbral). El método POT es la técnica más usada para analizar la cola de una función de distribución. Estas dos técnicas están basadas en modelos distribucionales obtenidas a partir de teorías asintóticas.

El problema a resolver es la estimación de los parámetros de estas distribuciones límites, en particular el parámetro de forma (también llamado el índice de valor extremo,  $\zeta$ ) el cual determina el comportamiento de los valores extremos. Determinar el umbral en el método POT, conlleva a un *trade-off* entre sesgo y varianza en la estimación de los parámetros de la función de distribución que se asume para ajustar los valores extremos. Al usar métodos basados en cuantiles (por ejemplo el estimador de Hill), también depende de la elección *apropiada* de estadísticos ordenados superiores. Elegir el umbral óptimo (en el método POT) conlleva al mismo problema de elegir el número óptimo  $k$  de estadísticos ordenados superiores (en el estimador de Hill). Si se utiliza el método *block maxima*, sesgo se presenta con bloques muy pequeños y varianza en el caso contrario.

De esta manera, la estimación del índice de valor extremo se convierte en un problema importante para estimar de manera confiable cuantiles altos como una medida de riesgo siguiendo la técnica EVT. Sin embargo, la selección de  $k$  (o del umbral) para estimar el índice de valor extremo no es una tarea fácil. El propósito de este documento, es entonces, revisar algunos métodos escogidos para estimar el IVE, con el fin de mitigar los problemas anteriormente mencionados.

El enfoque de EVT en este trabajo está principalmente basado en Beirlant et al. (2004) y el Capítulo 7 de McNeil et al. (2005). El documento está organizado como sigue: el Capítulo 1 describe brevemente el marco conceptual de EVT. El Capítulo 2 revisa los estimadores elegidos para modelos tipo-Pareto. El Capítulo 3 presenta y comenta los resultados de las simulaciones calculando los estimadores elegidos a varias funciones de distribución. En el Capítulo 4 se presentan los resultados de los métodos aplicados a un caso, "Danish Fire data". Finalmente el Capítulo 5 concluye el documento.

## Marco Conceptual

Esta sección está basada en el Capítulo 7 de McNeil et al. (2005), la cual introduce los principales métodos revisados en la literatura en estimación de cuantiles para datos que presentan distribuciones con colas espesas:

- Block Maxima
- Método POT

### 1.1 Block Maxima

El método de *block maxima* está basado en el comportamiento límite de los máximos normalizados de variables aleatorias independientes e idénticamente distribuidas (iid). La sucesión de estas variables aleatorias iid pueden ser pérdidas financieras o por siniestros en el área de seguros.

#### 1.1.1 Distribución de Valor Extremo Generalizada (GEV)

*Convergencia de sumas.* El rol de la distribución de valor extremo generalizada (GEV, por sus siglas en inglés) en teoría de extremos es análogo a la distribución normal en el teorema de límite central para sumas de variables aleatorias. Asumiendo que las variables aleatorias subyacentes  $X_1, X_2, \dots$  son variables aleatorias iid con una varianza finita y sea  $S_n = X_1 + \dots + X_n$  la suma de las primeras  $n$  variables aleatorias, la versión estándar del teorema del límite central se puede expresar como:

$$\lim_{n \rightarrow \infty} P\left(\frac{S_n - a_n}{b_n} \leq x\right) = \Phi\left(\frac{x}{\sigma}\right) \quad x \in \mathbb{R},$$

Donde  $\Phi$  denota la función de distribución normal estándar. La normalización apropiada usa sucesiones de constantes normalizadas  $(a_n)$  y  $(b_n)$  definidas por  $a_n = nE(X_1)$  y  $b_n = \sqrt{\text{var}(X_1)}$ .

*Convergencia de máximos.* La teoría del valor extremo (EVT, por sus siglas en inglés) se refiere a distribuciones límites para máximos normalizados  $M_n = \max(X_1, \dots, X_n)$  de variables aleatorias iid, denominados *block maxima* (máximos por bloque). Las únicas distribuciones límite no degenerativas posibles para máximos por bloque normalizados son la familia de valor extremo generalizado (GEV). La siguiente definición se puede encontrar en la Definición 7.1 de McNeil et al. (2005):

**Definición 1** (distribución GEV). La función de distribución de la distribución GEV estándar está dada por:

$$H_\xi(x) = \begin{cases} \exp\left(-\left(1 + \xi x\right)^{1/\xi}\right) & \text{si } \xi \neq 0 \\ \exp(-\exp(-x)) & \text{si } \xi = 0 \end{cases}$$

Donde  $1 + \xi x > 0$

Una familia de tres parámetros es obtenida definiendo  $H_{\xi, \mu, \sigma} := H_\xi((x - \mu) / \sigma)$  para un parámetro de localización  $\mu \in \mathbb{R}$  y un parámetro de escala  $\sigma > 0$ . El parámetro  $\xi$  es conocido como el parámetro de *forma* de la distribución GEV y  $H_\xi$  define un tipo de distribución, es decir, una familia de distribuciones específica dependiendo de los parámetros de localización y escala. Cuando  $\xi > 0$  la distribución GEV es una distribución Fréchet, cuando  $\xi = 0$  ésta es una distribución Gumbel, cuando  $\xi < 0$  ésta es una distribución Weibull.

La distribución Weibull es una distribución de cola corta con un *punto final a la derecha*. El punto final de una distribución se denota por  $x_F = \sup \{x \in \mathbb{R} : F(x) < 1\}$ . Por lo contrario, las distribuciones Gumbel y Fréchet tienen puntos finales a la derecha infinitos, pero el decaimiento de la cola de la distribución Fréchet es mucho más lento que el de la distribución Gumbel.

El rol de la distribución GEV en el estudio de máximos es formalizado por la siguiente definición y teorema:

**Definición 2** (dominio máximo de atracción). Si

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n - d_n}{c_n} \leq x\right) = \lim_{n \rightarrow \infty} F^n\left(\frac{x + d_n}{c_n}\right) = H(x)$$

se cumple para alguna función de distribución no degenerativa  $H$  y sucesiones de constantes reales  $(d_n)$  y  $(c_n)$ , entonces  $F$  pertenece al dominio máximo de atracción de  $H$  y se denota como  $F \in \text{MDA}(H)$

MDA representa las siglas en inglés de dominio máximo de atracción. La sucesión de constantes  $c_n > 0$  para todo valor de  $n$ . Y por notación  $P(M_n \leq x) = F^n(x)$ . El siguiente teorema corresponde al Teorema 7.3 de McNeil (2005).

**Teorema 3** (Fischer-Tippet, Gnedenko). Si  $F \in \text{MDA}(H)$  para alguna función de distribución no degenerativa  $H$ , entonces  $H$  debe ser una distribución del tipo  $H_\xi$ , es decir, una distribución GEV

Una versión más formal de este teorema y un bosquejo de la prueba se encuentra en el Teorema 3.2.3 de Embrechts et al. (1997). Más detalles se pueden encontrar en Resnick

(1987). McNeil et al. (2005) definen función de distribución no degenerativa como una distribución límite la cual no está concentrada en un punto en particular.

### 1.1.2 El Dominio Máximo de Atracción Fréchet

Todas las funciones de distribución continua comúnmente usadas en administración de riesgos pertenecen al  $MDA(H_\xi)$ , para algún valor de  $\xi$ . En esta sección consideraremos solo el caso Fréchet, es decir, cuando  $\xi > 0$ . Para los casos Gumbel y Weibull remitimos al lector a McNeil et al. (2005) y las referencias allí contenidas.

La siguiente definición y teorema corresponden a la Definición 7.7 (i) y Teorema 7.8 de McNeil et al. (2005) respectivamente:

**Definición 4** (función de variación lenta). Una función positiva, Lebesgue-medible  $L$  en  $(0, \infty)$  es de variación lenta en  $\infty$  si:

$$\lim_{u \rightarrow \infty} \frac{L(xu)}{L(x)} = 1, \quad x > 0. \square$$

Ejemplos de funciones de variación lenta están las constantes, logarítmicas, potencias de logaritmos y funciones de logaritmos iterados.

**Teorema 5** (MDA Fréchet, Gnedenko). Para  $\xi > 0$ ,

$$F \in MDA(H_\xi) \Leftrightarrow 1 - F(x) = x^{-1/\xi} L(x)$$

Para alguna función de variación lenta  $L$  en  $\infty$  □

La prueba de este teorema se puede encontrar en Embrechts et al. (1997). Estas distribuciones son las más estudiadas en EVT y en este documento nos enfocaremos en la estimación de  $\xi > 0$ .

Las colas de estas distribuciones decaen como una función de potencia y la tasa de decaimiento  $\alpha = 1/\xi$  es denominado el *índice de la cola* de la distribución.  $\xi$  se denomina el parámetro de *índice de valor extremo* (IVE).

### 1.1.3 Distribuciones tipo-Pareto (colas espesas)

Una distribución tipo-Pareto o de colas espesas es una distribución  $F$  la cual satisface:

$$(1.1) 1 - F(x) = x^{-1/\xi} L(x), \quad x \rightarrow \infty, \quad \xi > 0,$$

donde  $L$  es una función de variación lenta para todo  $x > 0$ . Distribuciones tipo-Pareto también se denominan “distribuciones con cola de Pareto”. Ejemplos de estas distribuciones son Pareto, gamma inversa,  $t$ -Student, loggamma,  $F$  y Burr.

#### 1.1.4 El método Block Maxima

Suponga que tenemos datos de una distribución subyacente  $F$  desconocida y que los datos pueden ser divididos en  $m$  bloques de tamaño  $n$ . Asumimos que la distribución extrema de nuestros datos, de hecho se distribuyen GEV con algunos parámetros desconocidos  $\zeta, \mu, \sigma$ . Podemos estimar estos parámetros usando máxima verosimilitud (las observaciones de los máximos por bloque se asumen ser independientes). Pero la selección del tamaño del bloque genera un *trade-off* entre sesgo y varianza. Bloques pequeños generan sesgo y bloques muy grandes generan alta varianza en la estimación de los parámetros. Vea el Ejemplo 7.12 de McNeil et al. (2005) de una aplicación del método block maxima a datos financieros.

### 1.2 Picos Sobre Umbral (POT)

El método POT por sus siglas en inglés, es útil para grandes observaciones las cuales exceden un umbral alto  $u$ . Este método es más útil que block maxima en aplicaciones prácticas, debido al uso más eficiente de los datos en valores extremos (McNeil (1999)).

#### 1.2.1 Distribución Generalizada de Pareto (GPD)

GPD es una función de distribución de dos parámetros con la siguiente representación:

$$G_{\zeta, \beta} = \begin{cases} 1 - \left(1 + \zeta x / \beta\right)^{-1/\zeta} & \text{si } \zeta \neq 0 \\ 1 - \exp(-x/\beta) & \text{si } \zeta = 0 \end{cases}$$

donde  $\beta > 0$ , y  $x \geq 0$  cuando  $\zeta \geq 0$ ; y  $0 \leq x \leq \beta/\zeta$  cuando  $\zeta < 0$ . Los parámetros  $\zeta$  y  $\beta$  son los parámetros de *forma* y *escala* respectivamente. Cuando  $\zeta > 0$ , la distribución  $G_{\zeta, \beta}$  es la distribución estricta de Pareto. Cuando  $\zeta = 0$  tenemos la distribución exponencial y cuando  $\zeta < 0$  tenemos la distribución de Pareto tipo II.

En términos de dominio de atracción tenemos que  $G_{\zeta, \beta} \in \text{MDA}(H_{\zeta})$  para todo  $\zeta \in \mathbb{R}$ . El caso cuando  $\zeta > 0$  corresponde a distribuciones de colas espesas y se puede verificar que los momentos  $E[X^j] = \infty$  para  $j \geq 1/\zeta$ .

Las siguientes definiciones y teorema corresponden a las Definiciones 7.17, 7.18 y Teorema 7.20 de McNeil (2005) respectivamente:

**Definición 6** (distribución de excesos sobre un umbral  $u$ ). Sea  $X$  una variable aleatoria con función de distribución  $F$ . La distribución de excesos sobre el umbral  $u$  tiene una función de distribución dada por:

$$F_u(x) = P(-u \leq x | X > u) = \frac{F(x+u) - F(u)}{1 - F(u)},$$

Para  $0 \leq x < x_F$ , donde  $x_F \leq \infty$  es el punto final a la derecha de  $F$ .

Esta distribución es también conocida como la distribución condicional de los excesos o la distribución de las pérdidas en exceso.

**Definición 7** (función de exceso medio). La función de exceso medio (mef) de una variable aleatoria  $X$  con una media finita está dada por:

$$e(u) = E(X - u | X > u)$$

La *mef* es también conocida como la función de vida residual media.

**Teorema 8** (Pickands-Balkema-de Haan). Podemos encontrar una función positiva-medible  $\beta(u)$  tal que:

$$\lim_{u \rightarrow x_F} \sup_{0 \leq x < x_F - u} |F_u(x) - G_{\xi, \beta}(x)| = 0,$$

sí y solo sí  $F \in \text{MDA}(H_\xi)$ ,  $\xi \in \mathbb{R}$ .

Detalles adicionales de este teorema se puede encontrar en Embrechts et al. (1997). El teorema dice que para una función de distribución subyacente  $F$  que está en el dominio de atracción de la distribución de valor extremo generalizada, cuando el umbral  $u$  se incrementa progresivamente, la distribución de excesos  $F_u$  converge a una distribución generalizada de Pareto (GPD). Además el parámetro de forma del modelo GPD para los excesos es el mismo parámetro de forma de la distribución GEV para los máximos.

Para poder aplicar el método POT se requiere del siguiente supuesto (ver Supuesto 7.21 de McNeil et al. (2005)):

Sea  $F$  una distribución de pérdidas con un punto final a la derecha  $x_F$  se asume que para algún umbral alto  $u$ , tenemos  $F_u(x) = G_{\xi, \beta}(x)$  para  $0 \leq x < x_F - u$  y algún  $\xi \in \mathbb{R}$  y  $\beta > 0$ . Esto es, asuma que  $F \in \text{MDA}(H_\xi)$ , así que para algún umbral alto  $u$  escogido adecuadamente, podemos modelar  $F_u$  mediante una GPD.

### 1.2.2 El método POT

Dado unos datos de pérdida  $X_1, \dots, X_n$  de una función de distribución (desconocida)  $F$ , un número aleatorio  $N_u$  de pérdidas excederá nuestro umbral  $u$ ; nombramos estos datos

como:  $\tilde{X}_1, \dots, \tilde{X}_{N_u}$ . Para cada uno de estos excesos calculamos la cantidad  $Y_j = \tilde{X}_j - u$  de las pérdidas en exceso. Deseamos estimar los parámetros de un modelo GPD ajustando esta distribución a los  $N_u$  pérdidas en exceso. Estimamos estos parámetros usando máxima verosimilitud (se asume que los datos en exceso son independientes).

Pero la selección del umbral genera un *trade-off* entre sesgo y varianza en la estimación. Valores muy bajos del umbral generan sesgo en la estimación, mientras que valores muy altos del umbral generan alta varianza en la estimación.

Vea el Ejemplo 7.23 de McNeil et al. (2005) de una aplicación del método POT a un caso de seguros, el caso “Danish fire loss data”, el cual retomaremos en el Capítulo 4 de este documento. El Ejemplo 7.24 de McNeil et al. (2005) aplica POT a datos financieros. En el caso Danish data calcularemos también altos cuantiles como medidas de riesgo. Las siguientes definiciones corresponden a la definición 2.10 y 2.15 de McNeil et al. (2005):

**Definición 9 (Value at Risk - VaR).** Dado algún nivel de confiabilidad  $\alpha \in (0, 1)$ . El VaR de un portafolio al nivel de confiabilidad  $\alpha$  está dado por el número más pequeño  $l$ , tal que la probabilidad de que la pérdida  $L$  exceda  $l$  no es más grande que  $(1 - \alpha)$ . Formalmente,

$$\text{VaR}_\alpha = \inf \{ l \in \mathbb{R} : P(L > l) \leq (1 - \alpha) \} = \inf \{ l \in \mathbb{R} : F_L(l) \geq \alpha \}.$$

En términos probabilísticos, VaR es simplemente un cuantil de la distribución de pérdidas.

**Definición 10 (Expected Shortfall – ES).** Para una pérdida  $L$  con  $E(|L|) < \infty$  con función de distribución  $F_L$ , el expected shortfall a un nivel de confiabilidad  $\alpha \in (0, 1)$  se define como:

$$ES_\alpha = \frac{1}{1 - \alpha} \int_\alpha^1 q_u(F_L) du,$$

Donde  $q_u(F_L) = F_L^{-1}(u)$  es la función cuantil de  $F_L$ .

En otras palabras, para funciones de distribución continuas, el expected shortfall es el valor esperado de las pérdidas dado que éstas superaron un cuantil a nivel  $\alpha$  (o el  $\text{VaR}_\alpha$ ).



## Estimación del Índice de Valor Extremo (IVE) $\zeta$

El método GPD no es la única aproximación para estimar los parámetros, en específico el parámetro de forma, de la cola de una distribución. Otro método es el conocido enfoque de Hill que se describe en la próxima sección.

### 2.1 El estimador de Hill

El *estimador de Hill* es el estimador más popular de índice de valor extremo (IVE) el cual está restringido al caso cuando  $\zeta > 0$ . Es decir, que es aplicable para distribuciones que pertenecen al MDA de Fréchet (ver Teorema 5 de este documento).

Dado estadísticos ordenados  $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$  el estimador de Hill toma la siguiente forma:

$$(2.1) \hat{\zeta}_{k,n} = \frac{1}{k} \sum_{j=1}^k \ln X_{n-j+1,n} - \ln X_{n-k,n}, \quad k = 1, \dots, n-1,$$

donde  $X_{n-k,n}$  puede ser visto como el umbral. Los valores por encima del umbral se denotan como  $X_{n-j+1,n}$  y  $j$  toma valores  $1, \dots, k$  ( $k = 1, \dots, n-1$ ).

Embrechts et al. (1997) muestran varios métodos para obtener el estimador de Hill. Beirlant et al. (2004), también obtienen el estimador de Hill mediante punto de vista de cuantiles, punto de vista de probabilidad y otros enfoques.

Por lo general se usa el gráfico de Hill para estimar  $\zeta$ , el cual es un gráfico de  $\hat{\zeta}_{k,n}$  versus  $k$ , y se escoge el estimado del índice de valor extremo donde el gráfico es estable para algún valor de  $k$  no muy pequeño ni muy grande. Usualmente vemos alta variabilidad en el gráfico para valores pequeños de  $k$ . Esto se debe a que hay muy pocas observaciones de datos extremos (pérdidas) y gran diferencia entre ellos. Si escogemos un valor grande de  $k$ , donde se muestran casi todos los datos, se puede notar sesgo en el gráfico de Hill.

La siguiente Figura muestra un ejemplo del gráfico de Hill para 5000 datos iid de una distribución de Pareto ( $\alpha = 1$ ) con  $\zeta = 1$ .

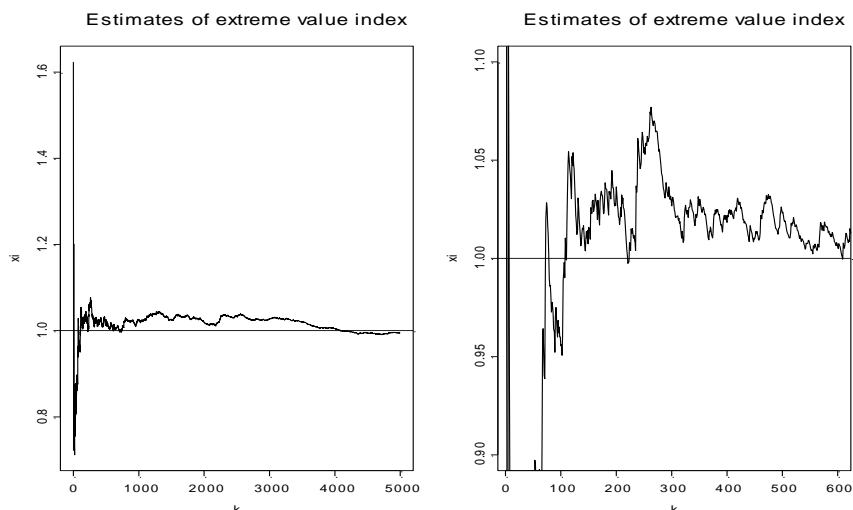


Figura 1. Gráfico de Hill para 5000 datos iid (izquierda) de una distribución de Pareto ( $\alpha = 1$ ) con  $\zeta = 1$ ; y su versión expandida (derecha) hasta 600 estadísticos ordenados. La línea horizontal muestra el verdadero valor de  $\zeta$ .

El estimador de Hill parece dar una buena aproximación del verdadero valor de  $\zeta$  para el ejemplo anterior.

De acuerdo a Drees et al. (2000):

“In fact,...the traditional Hill plot is most effective only when the underlying distribution is Pareto or very close to Pareto. For the Pareto distribution one expects the Hill plot to be close to  $\gamma$  for the right side of the plot, since the Hill estimator is the maximum likelihood estimator in the Pareto model.”

En su artículo, ellos denotan el IVE como  $\gamma$ . Esta conclusión se puede observar en la Figura 1. De hecho para datos iid, el estimador de Hill es (débilmente) consistente<sup>1</sup> para  $\zeta$ .

Si  $k, n \rightarrow \infty$ , y  $k/n \rightarrow 0$ , entonces:

$$\hat{\zeta}_{k,n}^{(H)} \xrightarrow{P} \zeta$$

Pero, funciona el estimador de Hill en la realidad? La siguiente Figura es tomada del artículo de Matthys y Beirlant (2000); y representa los estimados mediante el enfoque de Hill para caídas en porcentaje del S&P 500. El periodo de observación es Enero de 1980 a Octubre 16 de 1987.

<sup>1</sup>Ver Embrechts et al. (1997) para más detalles de propiedades del estimador de Hill.

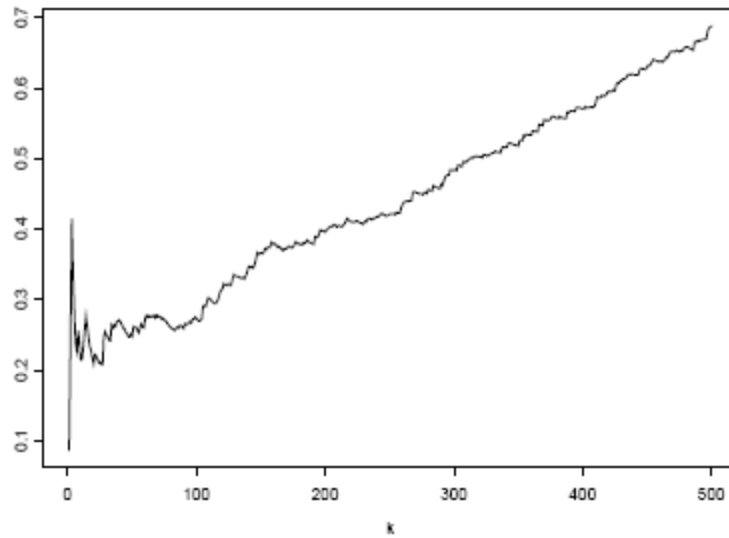
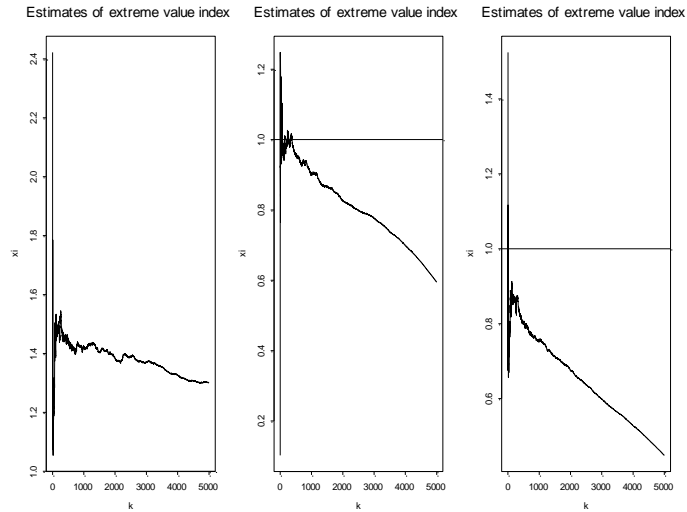


Figura 2. Gráfico de Hill para las caídas del S&P 500. Fuente: Matthys y Beirlant (2000).

Para este ejemplo real no es claro que valor de  $k$  escoger para estimar el índice de valor extremo. Para pequeños valores de  $k$  notamos gran varianza del estimador, pero si  $k$  es relativamente alto, observamos un sesgo significativo en la estimación del IVE.

La siguiente Figura muestra otros ejemplos con funciones conocidas:



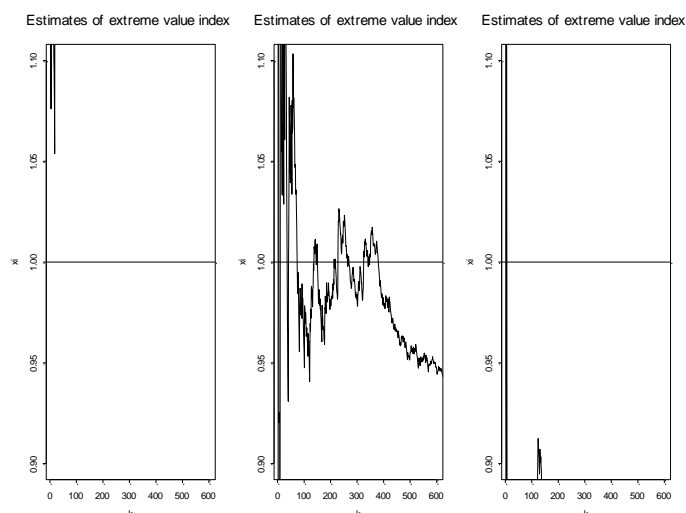


Figura 3. Gráficos de Hill para 5000 datos iid de algunas distribuciones tipo-Pareto con IVE = 1 (arriba); y su versión expandida (abajo). La línea horizontal indica el verdadero valor de IVE.

En los anteriores ejemplos los gráficos de Hill no son útiles para escoger  $k$  y por consiguiente no podemos tener un estimado confiable de  $\zeta$ . Los datos en la Figura 3 corresponden a las siguientes funciones de supervivencia ( $1 - F(x)$ ):

- $x^{-1/\zeta} (1 + x^{-0.5})$  (izquierda)
- $x^{-1/\zeta} (1 / \log x)$  (centro)
- $x^{-1/\zeta} (1 / \sqrt{\log x})$  (derecha)

La función de supervivencia de la distribución de Pareto es  $x^{-1/\zeta}$ , donde la función de variación lenta  $L$ , es igual a 1 (ver sección 1.1.3 de este documento). El problema central radica en que una distribución tipo-Pareto no significa que la cola es exactamente como la de una distribución de Pareto, debido a que la función de variación lenta no es constante.

Como se mencionó al comienzo del documento el problema de escoger  $k$  en el gráfico de Hill es el mismo problema de escoger el umbral  $u$  para el método POT. Con respecto a la selección del umbral, Embrechts y Nešlehová (Notas de clase) concluyen:

- la tasa de convergencia depende de  $F$ , en particular de la función de variación lenta  $L$ , en una manera complicada y puede ser muy lenta (este punto se explicará más abajo cuando hablemos del parámetro de segundo orden).
- $L$  no es visible directamente de los datos.
- La selección del umbral es muy difícil. Trade-off entre sesgo y varianza se da por lo general.

Lo que se observa no solo en la Figura 3 (donde se conoce  $L$ ) pero también en la Figura 2 (donde no se conoce  $L$ ) para el caso de las caídas del S&P.

Entonces, cómo solucionar el problema de trade-off entre sesgo y varianza? Algunos autores han propuesto suavizar el gráfico de Hill. La idea es mitigar el problema de alta volatilidad observada en el gráfico de Hill (para valores relativamente pequeños de  $k$ ).

Bajo condiciones de variación regular de segundo orden, el estimador de Hill es asintóticamente normal con varianza asintótica  $\zeta^2$  para datos iid. Esta condición dice:

Para  $\zeta > 0, \rho \leq 0, c \in \mathbb{R}$

$$\lim_{t \rightarrow \infty} \frac{\frac{L(x) - 1}{L(x)} - 1}{A(x)} = c \frac{x^\rho - 1}{\rho},$$

Donde  $A(t) \rightarrow 0, t \rightarrow \infty$ . Para más detalles, refiérase a Embrechts et al. (1997) y las referencias allí contenidas.

El parámetro  $\rho$  es denominado *el parámetro de segundo orden* y mide la velocidad de convergencia en EVT. La condición de variación regular de segundo orden tiene dos objetivos: (1) establecer la normalidad asintótica de los estimadores de IVE y (2) estudiar las tasas de convergencia a distribuciones de valor extremo. Como veremos más adelante,  $\rho$  juega un papel importante en la estimación de  $\zeta$ .

Otros autores han intentado reducir el nivel de sesgo observado en el gráfico de Hill. Nuestra primera revisión es el estimador de average Hill propuesto por Resnick y Stărică (1997), quienes reducen la varianza asintótica del estimador de Hill, pero no sesgo. Entonces revisaremos también el método de Zipf, seguido por una técnica de reducción de sesgo (modelo EGPD) pero que aumenta en varianza con respecto al estimador de Hill.

Por lo tanto, cuál es el mejor método a usar para estimar el IVE si existe? Esta es la pregunta que trataremos de resolver en el todo el documento. A continuación se presenta una breve descripción de tres métodos alternativos para la estimación del índice de valor extremo (IVE).

## 2.2 El estimador average Hill

Un método para resolver el problema de alta variabilidad en el gráfico de Hill es propuesto por Resnick y Stărică (1997). La idea consiste en promediar los valores del estimador de Hill correspondiente a diferentes valores  $p$  de los estadísticos ordenados. Los estadísticos ordenados son los datos de una muestra aleatoria ordenados de menor a mayor.

El modelo es:

$$\hat{\xi}_{k,n,c}^{(vH)} = \frac{1}{c-1} \sum_{p=k+1}^{ck} \hat{\xi}_{p,n}^{(vH)}, \quad k = 1, \dots, n-1,$$

Donde  $c > 1$ , y  $\hat{\xi}_{p,n}^{(vH)}$  corresponde al estimador de Hill. El estimador average Hill funciona bien cuando los datos representados en el gráfico de Hill exhiben alta volatilidad pero no cuando se presenta alto sesgo, como se observa en el panel derecho la siguiente Figura.

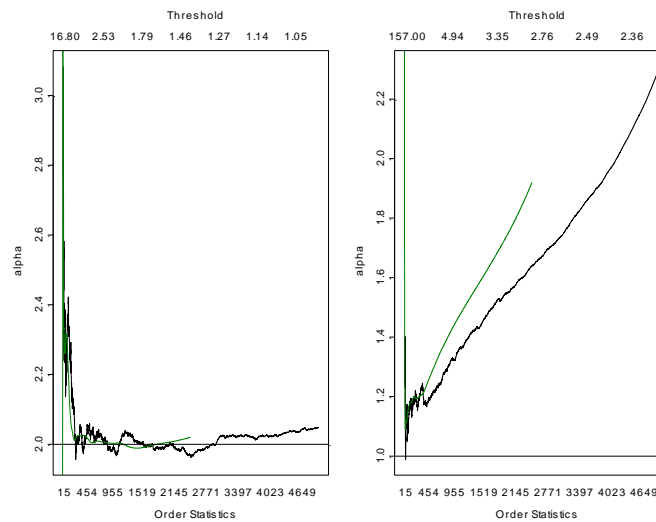


Figura 4. Estimador de Hill (línea negra) y average Hill (línea verde) para 5000 datos iid de una distribución Pareto ( $1 - F(x) = x^{-1/\xi}$ ) con  $\alpha = 2$  (izquierda), y para  $1 - F(x) = x^{-1/\xi} (1 / \sqrt{\log x})$  con  $\alpha = 1$  (derecha);  $\alpha = 1/\xi$  y  $c = 2$  para ambos casos. La línea horizontal representa el verdadero valor de  $\alpha$ .

La variable  $c$  puede ser interpretado como el parámetro promediador de los valores del estimador de Hill. Un resultado importante es que la varianza del estimador average Hill decrece con el valor de  $c$ . Adicionalmente, existe una reducción de la varianza con respecto al estimador de Hill, por ejemplo si  $c = 2$ , la varianza del estimador average Hill es  $0.653\xi^2$  (puesto que la varianza está dado por  $2\xi^2/c [1 - (\log c)/c]$ ) es decir, una reducción del 34.7% de la varianza con respecto al del estimador de Hill. Por consiguiente, se desearía escoger un valor de  $c$  tan grande como sea posible para reducir la varianza de este estimador. Sin embargo al incrementar el valor de  $c$  se obtiene una reducción de datos a ser graficados dificultando así la selección del valor verdadero del IVE. Como una solución a este problema, Resnick y Stărică proponen utilizar  $c$  entre  $n^{0.1}$  y  $n^{0.2}$ , donde  $n$  es el tamaño de los datos.

## 2.3 El estimador de Zipf

Una herramienta gráfica muy útil para un análisis descriptivo de datos es el QQ-plot (Quantile-Quantile plot). Esta herramienta es útil para determinar si una serie de datos de tamaño  $n$  proviene de una población con una distribución paramétrica específica. Si los cuantiles de una distribución paramétrica están linealmente relacionados con los cuantiles de los datos de una muestra en el QQ-plot, se obtiene un ajuste de los datos de la muestra a dicha distribución paramétrica. La distribución normal se ha empleado en muchas aplicaciones estadísticas. Pero los datos financieros (y de seguros) por lo general exhiben colas espesas que no son adecuadamente modelados por una distribución normal. Para este tipo de datos es mejor usar el Pareto-Quantile plot (también denominado el Zipf plot) que se construye utilizando los siguientes valores para los ejes  $\{x,y\}$ :

$$(2.2) \quad \left\{ \log\left(\frac{n+1}{j}\right), \log X_{n-j+1} \right\}, \quad j = 1, \dots, n.$$

Si los datos ( $X_j$ ) siguen una distribución estricta de Pareto, la curva obtenida siguiendo (2.2), en el Pareto-Quantile plot se aproxima a una línea recta. Además su pendiente es aproximadamente igual a  $\zeta$ . Si los datos siguen (1.1) la curva graficada en el Pareto-Quantile plot es aproximadamente lineal pero para valores pequeños de  $j$ .

La pendiente de la línea resultante de aplicar mínimos cuadrados a los puntos obtenidos de (2.2) es un estimador denominado el “estimador-qq” (Kratz y Resnick (1996)). Puesto que Zipf usó este estimador desde finales de la década de los 20’s, el estimador-qq también es conocido como el “estimador de Zipf”. Schultze y Steinebach (1996) también propusieron el mismo estimador usando un procedimiento similar (ver Csörgő y Viharos (1998)). El modelo es:

$$\hat{\zeta}_{k,n} = \frac{\sum_{j=1}^k \log\left(\frac{k+1}{j}\right) \log X_{n-j+1,n} - \frac{1}{k} \sum_{j=1}^k \log\left(\frac{k+1}{j}\right) \sum_{j=1}^k \log X_{n-j+1,n}}{\sum_{j=1}^k \left( \log^2\left(\frac{k+1}{j}\right) \right) - \frac{1}{k} \left( \sum_{j=1}^k \log\left(\frac{k+1}{j}\right) \right)^2}, \quad k = 1, \dots, n-1.$$

El estimador de Zipf es consistente si  $k \rightarrow \infty$  y  $k/n \rightarrow 0$ . Bajo otras condiciones de segundo orden en  $F(x)$  y restricciones sobre  $k(n)$ , este estimador es asintóticamente normal con media asintótica  $\zeta$  y varianza asintótica  $2\zeta^2$ , (ver Kratz y Resnick (1996)).

Por consiguiente el estimador de Hill presenta menor varianza que el estimador de Zipf. Sin embargo una ventaja del estimador de Zipf con respecto al estimador de Hill es que los residuos de la curva obtenida en el Pareto-Quantile plot contienen información que potencialmente podría ser usado para mitigar el sesgo en los estimados cuando las colas de la distribución empírica de los datos no siguen una distribución Pareto.

La siguiente Figura muestra una comparación del estimador de Zipf y Hill para datos simulados de una distribución Pareto:

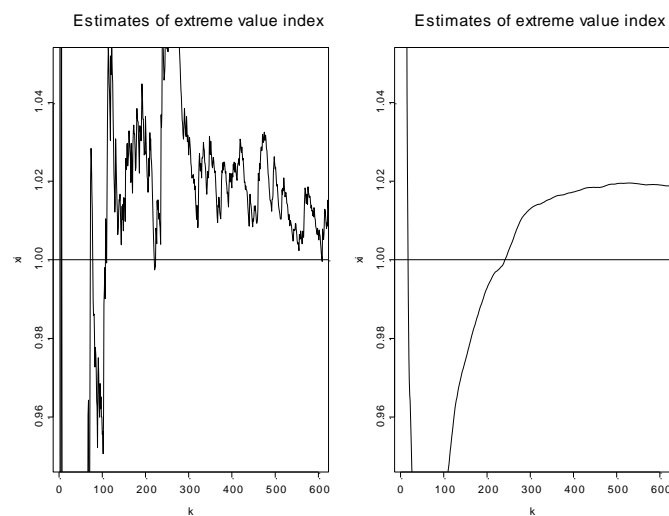


Figura 5. Estimador de Hill (izquierda) y estimador de Zipf (derecha) para 5000 datos iid de una distribución Pareto ( $\alpha = 1$ ). La línea horizontal representa el verdadero valor de  $\zeta$ .

## 2.4 El método de la Distribución Generalizada de Pareto extendido (EGPD)

La definición de una distribución tipo-Pareto (ver sección 1.1.3 de este documento) puede ser re-escrita como:

$$\lim_{u \rightarrow \infty} \frac{1 - F(x)}{1 - F(u)} = x^{-1/\xi},$$

Para cualquier  $x > 1$ , y esta puede ser interpretada como:



$$(2.3) \quad P\left\{\frac{X}{u} > u \mid X > u\right\} \approx x^{-1/\xi},$$

Para valores grandes de  $u$  y  $x > 1$ .

La relación (2.3) dice que podemos aproximar la distribución de los excesos relativos condicionados en  $X_i > u$  (de ahora en adelante, la distribución condicional de los excesos relativos) a una distribución estricta de Pareto.

La distribución condicional de los excesos relativos en el caso de las distribuciones tipo-Pareto satisface:

$$(2.4) \quad P\left\{\frac{X}{u} > u \mid X > u\right\} = x^{-1/\xi} \frac{L(x)}{L(u)}$$

Para todo  $x \geq 1$  y  $L$  siendo una función de variación lenta. Puesto que  $L(x)/L(u) \rightarrow 1$  para todo  $x > 0$  (ver Definición 4 de este documento), obtenemos (2.3) a partir de (2.4).

**Condición 10.** Existe una constante real  $\rho < 0$  y una función  $b$  que satisface  $b(u) \rightarrow 0$ , cuando  $u \rightarrow \infty$ , tal que para todo  $x \geq 1$ ,

$$(2.5) \quad \frac{\frac{L(x)}{L(u)} - 1}{b(u)} \rightarrow \frac{x^\rho - 1}{\rho}, \quad \text{cuando } u \rightarrow \infty$$

Para más detalles vea la Sección 3.12.1 de Bingham et al. (1987) □

Reemplazando el término  $\frac{L(x)}{L(u)}$  de (2.5) en (2.4) obtenemos:

$$P\left\{\frac{X}{u} > u \mid X > u\right\} = x^{-1/\xi} \left( 1 + b(u) \frac{x^\rho - 1}{\rho} + o(b(u)) \right), \quad \text{cuando } u \rightarrow \infty.$$

Eliminando el término remanente, es decir  $o(b(u))$ , y estableciendo  $\delta = b(u) / \rho$ , obtenemos

$$(2.6) P(X/u > u | X > u) = F_{\xi, \delta, \rho} = (1 - \delta) \left( \frac{x}{u} \right)^{-1/\xi} + \delta \left( \frac{x}{u} \right)^{-1/\xi + \rho}.$$

El rango de los parámetros es  $1/\xi\rho \leq \delta < 1$ ,  $\xi > 0$  y  $\rho < 0$ . La ecuación (2.6) puede ser visto como una mezcla de dos distribuciones de Pareto donde el parámetro ponderador  $\delta$  puede ser negativo (puesto que  $\rho$  es negativo).

Cuando  $\delta = 0$ , se obtiene el modelo estricto de Pareto para los excesos relativos. Cuando  $\rho = -1$  la distribución condicional de los excesos relativos asintóticamente equivale al modelo de distribución generalizada de Pareto (GPD).

De esta mezcla de dos distribuciones de Pareto se obtiene un nuevo estimador para el IVE mediante máxima verosimilitud. Esta nueva aproximación de la mezcla de distribuciones se denomina el “modelo de distribución de Pareto generalizada extendida” (EGPD).

Tanto  $\xi$  como  $\delta$  se estiman mediante máxima verosimilitud con una previa estimación de  $\rho$ . Se refiere al lector al Remark 3.2 de Beirlant et al. (2005) y las referencias allí contenidas para detalles de estimación externa de  $\rho$ .

Este modelo es difícilmente identificable cuando  $\rho$  es muy cercano a cero. La varianza asintótica de este nuevo estimador es igual a:

$$\left( \frac{1 - \rho}{\rho} \right)^2 \frac{\xi^2}{k}.$$

La siguiente Figura compara los estimados de Hill y modelo EGPD para datos simulados de una distribución  $t$  con tres grados de libertad ( $\rho = -2/3$ ). El gráfico para el nuevo estimador luce más estable alrededor del verdadero valor de  $\xi$  que para el estimador de Hill en este caso.

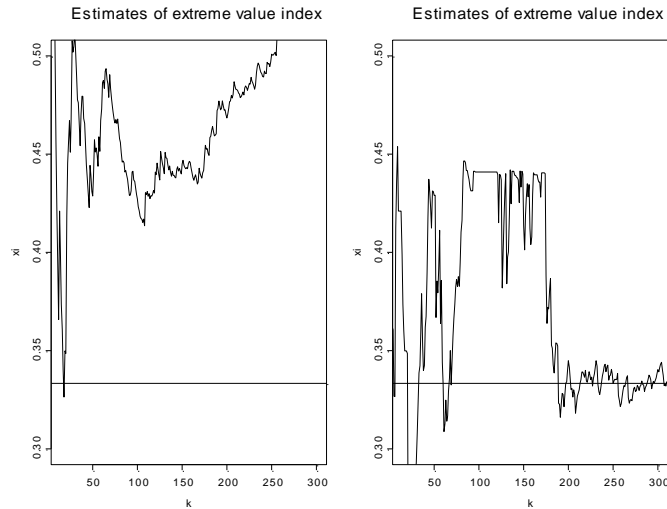


Figura 6. Estimador de Hill (izquierda) y modelo EGPD (derecha) para 1000 datos iid de una distribución  $t$ -Student con  $\nu = 3$ . La línea horizontal representa el verdadero valor de  $\zeta$ .

## 2.5 La selección de $k$

El gráfico  $\{k, \hat{\xi}_{k,n}\}$  es una herramienta adecuada para escoger  $k$  y así decidir el estimado de  $\zeta$ . Sin embargo, la selección óptima de  $k$  para escoger el verdadero valor del IVE no es una tarea fácil como se vio anteriormente.

Existen varios métodos adaptativos para escoger  $k$ ; vea por ejemplo la Sección 4.7 de Beirlant et al. (2004) y las referencias allí contenidas. Un método puede ser minimizar el *error cuadrático medio asintótico* (AMSE, por sus siglas en inglés) del estimador de Hill, que está dado por:

$$\begin{aligned} AMSE(\hat{\xi}_{k,n}) &= A \text{var}(\hat{\xi}_{k,n}) + ABias(\hat{\xi}_{k,n}) \\ &= \frac{\zeta^2}{k} + \left(\frac{b_{n,k}}{1-\rho}\right)^2. \end{aligned}$$

La idea es estimar  $\zeta$  mediante el estimador de Hill (vea por ejemplo la Sección 4 de Beirlant et al. (1999)), donde:

$$\hat{k} = \arg \min_{k; k > 10} \left( \frac{\hat{\xi}_k^2}{k} + \left( \frac{\hat{b}_{n,k}}{1 - \hat{\rho}_k} \right)^2 \right),$$

$\hat{\xi}$ ,  $\hat{b}$  y  $\hat{\rho}$  denotan los estimadores de máxima verosimilitud de  $\xi$ ,  $b$  y  $\rho$  respectivamente, posiblemente obtenidos mediante EGPD o por otro método alternativo.

En el Capítulo 4, aplicaremos los métodos para estimar el IVE al caso de “Danish fire data”. Seguiremos la idea de Reiss y Thomas (1997) para escoger  $k$  gráficamente. Para un número  $k$  intermedio, existe un balance entre varianza y sesgo del estimador, entonces un *plateau* (o meseta) se vuelve visible en el gráfico de Hill. Entonces buscaremos una región estable en los gráficos aplicando los diferentes métodos para estimar el IVE y decidir cuál es el valor de  $\xi$  en tal plateau.

En el siguiente capítulo realizaremos una simulación para probar los métodos de estimación del IVE a funciones de distribución comúnmente usadas en administración de riesgos para simular datos de pérdidas.

## Simulación

En esta sección comparamos el desempeño de los métodos anteriormente revisados usando simulación Monte Carlo. Los estimadores usados en la simulación son:

- El estimador average Hill,
- El estimador de Zipf,
- El método EGPD.

Se escogieron cuatro tipos de datos con distribuciones donde  $\xi > 0$  y dos tipos de datos con distribuciones donde  $\xi = 0$ ; ver Tabla 3.1.

También podemos clasificar las distribuciones de acuerdo al parámetro de segundo orden  $\rho$  (parámetro que mide la velocidad de convergencia en EVT):

- “Casos complicados” donde  $\rho \in (-1, 0)$ ; (distribución  $t$ -Student con  $\nu = 3, 4, 8$ ),
- “Casos *muy* complicados” donde  $\rho = 0$ ; (distribuciones lognormal, Weibull y loggamma),
- “Casos normales” donde  $\rho < -1$ ; (distribuciones Cauchy estándar y Pareto).

Esta clasificación corresponde a los casos en que el estimador de Hill puede funcionar bien o no y deseamos saber cuál es el comportamiento de los tres estimadores alternativos.

Tabla 3.1. Distribuciones usadas en la simulación

Distribuciones	Parámetros	EVI, $\xi$	$\rho$
$t$ - Student	$\nu = 3, 4, 8$	1/3, 1/4, 1/8	-2/3, -1/2, -1/4
Lognormal	$(\mu, \sigma) = (0, 1)$	0	0
Weibull	$\tau = 0.5$	0	0
Loggamma	$(\alpha, \beta) = (1, 2); (1, 10)$	1; 1	0
Cauchy estándar		1	-2
Pareto	$\alpha = 1/\xi$	1.2, 1.1, 1.0, 0.7, 0.5, 0.3	$-\infty$

Para cada una de estas distribuciones, se generaron 1000 muestras de tamaño 1000. El desempeño de cada estimador se evalúa en términos de la raíz del error cuadrático medio relativo (RRMSE, por sus siglas en inglés) basado en los  $k$  estadísticos ordenados

superiores para la mayoría de las distribuciones. Para las distribuciones lognormal y Weibull se usa el error cuadrático medio (MSE, por sus siglas en inglés).

$$RRMSE(\hat{\xi}) = \frac{\sqrt{MSE(\hat{\xi})}}{\xi}$$

Donde  $MSE(\hat{\xi}) = E(\hat{\xi} - \xi)^2$ . El verdadero valor de  $\xi$  se obtiene de la tercera columna de la Tabla 3.1. mientras que el valor esperado del estimador se obtiene de calcular la media de los estimados que arroje cada método en la simulación Monte Carlo para cada tipo de distribución.

Los siguientes resultados se obtuvieron utilizando los scripts en S-PLUS elaborados por Beirlant J., Goegebeur Y., Segers J., Teugels J., de Waal D., Ferro C., y Vandewalle B.; y se pueden encontrar en [ucs.kuleuven.be/Wiley/index.html](http://ucs.kuleuven.be/Wiley/index.html) (Chapter 4). Los archivos que se utilizaron fueron: Zipf.SSC, Hill2oPV.SSC para los métodos Zipf y EGDP respectivamente. La selección de los valores de los parámetros y algunas características de las funciones de distribución están principalmente basadas en McNeil y Saladin (1997), quienes realizaron un estudio similar aplicando el método POT.

### 3.1 La distribución $t$ – Student

La distribución  $t$  – Student ha sido empleado en el estudio de administración de riesgos, en especial para ajustar datos empíricos de retornos de títulos de renta variable para el cálculo de riesgo de mercado, puesto que la forma de la distribución  $t$  es similar a la distribución normal estándar. Recordemos que la distribución normal estándar fue originalmente usada para estimación de VaR (Valor en Riesgo), por su facilidad de cálculo y tratamiento. Aunque la distribución  $t$  es simétrica, ésta tiene colas más amplias que la normal, lo que permite ajustar mejor datos extremos que la normal.

La distribución  $t$  – Student  $F_\nu(x)$  tiene la siguiente función de densidad:

$$f_\alpha \propto \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

Donde  $\nu > 0$  se conoce como los grados de libertad. El caso cuando  $\nu = 1$  se refiere a la distribución Cauchy. Esta distribución pertenece al MDA ( $H_{1/\nu}$ ) y para valores grandes de  $x$  se puede mostrar que  $1 - F_\nu(x) \sim cx^{-\nu}$ , para alguna constante  $c$ . Una característica importante es que a medida que el número de grados de libertad tiende a infinito, la forma de la distribución  $t$  se aproxima a la normal estándar. Por tal razón hemos usado  $\nu = 3, 4$  y  $8$  y observar qué método se desempeña mejor cuando el número de grados de libertad es bajo.

Para la distribución  $t$  es posible mostrar que  $\lim_{\alpha \rightarrow 1} ES_{\alpha} / VaR_{\alpha} = \nu / (\nu - 1) > 1$ ; mientras que para la distribución normal tenemos  $\lim_{\alpha \rightarrow 1} ES_{\alpha} / VaR_{\alpha} = 1$ . Esto muestra que para una distribución con colas espesas la diferencia entre ES y VaR es más pronunciada que para la distribución normal (Ver Remark 2.22 de McNeil et al. (2005)).

Para este caso el estimador average Hill tiende a ser, en general, preferido para valores relativamente pequeños de  $k$ . Cuando se incrementa el número de grados de libertad, el método average Hill es mejor comparado con EGPD, pero el rango de estadísticos ordenados para su optimalidad decrece ( $k \leq 51$  para  $\nu = 3$ , y  $k \leq 32$  para  $\nu = 8$ ). Sin embargo el método EGPD tiene el mínimo RRMSE para  $\nu = 3, 4$  (9.8% en  $k = 160$  para  $\nu = 3$ , y 16.5% en  $k = 102$  para  $\nu = 4$ ). El estimador average Hill tiene el mínimo RRMSE (33.2%) para  $\nu = 8$  (pero  $k = 7$ ). El estimador de Zipf no es muy útil comparado con los otros dos métodos.

$t, \nu = 3 (\zeta = 1/3, \rho = -2/3)$

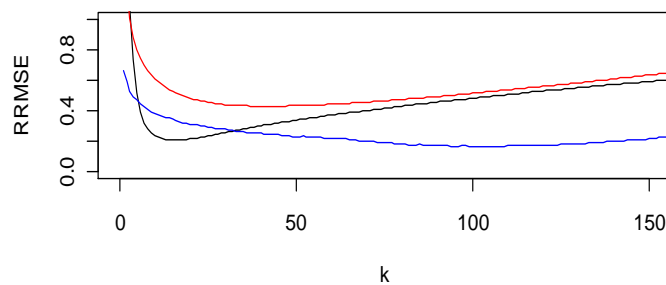
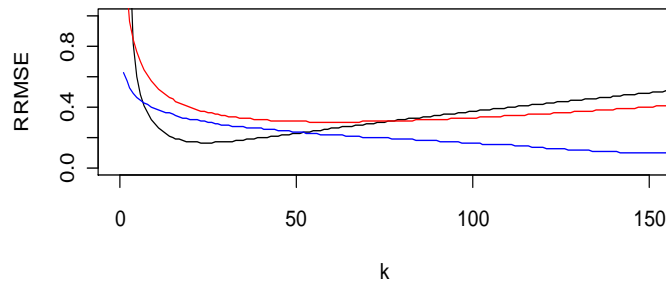
Método	RRMSE mín.
EGPD	0.098
Average Hill	0.167
Zipf	0.303

$t, \nu = 4 (\zeta = 1/4, \rho = -1/2)$

Método	RRMSE mín.
EGPD	0.165
Average Hill	0.206
Zipf	0.430

$t, \nu = 8 (\zeta = 1/8, \rho = -1/4)$

Método	RRMSE mín.
Average Hill	0.332
EGPD	0.492
Zipf	0.952



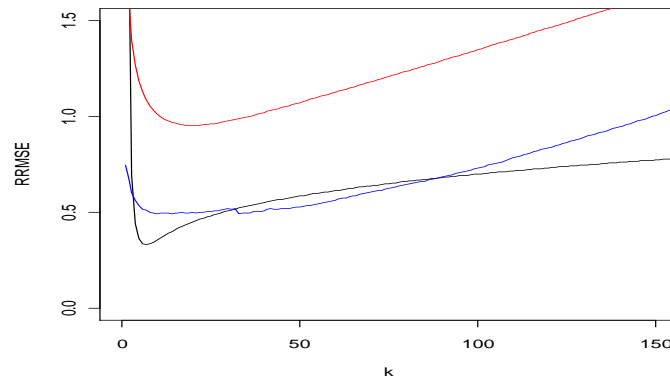


Figura 7. RRMSE para  $t$ -Student con  $v = 3$  (arriba), 4 (centro) y 8 (abajo). Método average Hill (línea negra), Zipf (línea roja) y EGPD (línea azul).

### 3.2. La distribución lognormal

La distribución lognormal ha sido utilizada para modelar datos de severidad en el sector de seguros o en riesgo operacional. La distribución lognormal también se ha utilizado como supuesto para modelar el precio de títulos de renta variable. Por ejemplo, el modelo de Black-Scholes usado para modelar el precio de opciones, supone que el precio del activo financiero subyacente sigue una distribución lognormal, es decir, que los retornos logarítmicos de este activo se distribuyen normalmente (una variable aleatoria  $X$  se distribuye lognormal si su logaritmo natural,  $Y = \log X$  se distribuye normal). La distribución lognormal tiene una forma asimétrica y toma solo valores positivos, esta última característica le da una ventaja a la distribución lognormal sobre la normal para modelar precios de acciones, puesto que una variable distribuida normalmente puede tomar valores positivos o negativos. Sin embargo varios estudios empíricos han mostrado que la distribución de retornos de activos financieros, por lo general no tiene una distribución normal, sino una distribución asimétrica con colas espesas. La distribución lognormal pertenece al  $MDA(H_0)$ , es decir que la función de distribución lognormal tiene un IVE igual a cero (ver Embrecht et al. (1997), para pruebas), y técnicamente tiene colas semi-espesas.

El método average Hill no se desempeña tan bien y no se incluye en la Figura 8; de acuerdo a las simulaciones, éste es el método con mayor MSE. Para todos los valores de  $k$  el modelo EGPD es el que mejor se desempeña.

Lognormal ( $\xi = 0, \rho = 0$ )

Método	MSE mín.
EGPD	0.057
Zipf	0.161
Average Hill	2.286



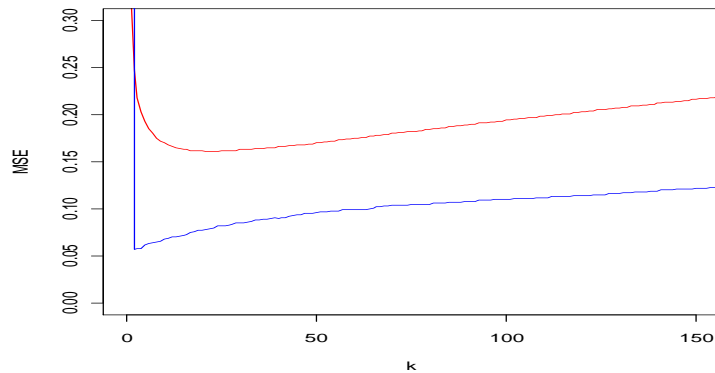


Figura 8. MSE para distribución lognormal. Método Zipf (línea roja) y EGPD (línea azul).

### 3.3 La distribución Weibull

Inicialmente su principal uso fue en la ciencia de materiales para simular el tiempo de vida de objetos, para cuantificar datos de fatiga; pero también se usa para modelar datos de severidades. La distribución Weibull es una distribución de dos parámetros y es conocida también como distribución *Extreme Value* Tipo III y está definida para valores positivos. Cuando el parámetro de forma es igual a 1, esta distribución se reduce a la exponencial, y cuando es igual a 2 se aproxima a la distribución Rayleigh (principalmente usada en aplicaciones de telecomunicaciones). Cuando el parámetro de forma es igual a 3.5 la distribución Weibull se aproxima a una normal. Es una *stretched-exponential* cuando el parámetro de forma es menor que uno y decae más lentamente que una exponencial. Por tal razón estamos interesados en estudiar el caso que su parámetro de forma,  $\tau < 1$ .

La distribución Weibull tiene también un IVE igual a cero. De nuevo el método de average Hill es el que tiene mayor MSE y no se incluye en la Figura 9. El método Zipf se desempeña mejor que el método EGPD para valores de  $k \leq 49$ .

Weibull ( $\zeta = 0, \rho = 0$ )

Método	MSE mín.
EGPD	0.165
Zipf	0.178
Average Hill	0.788

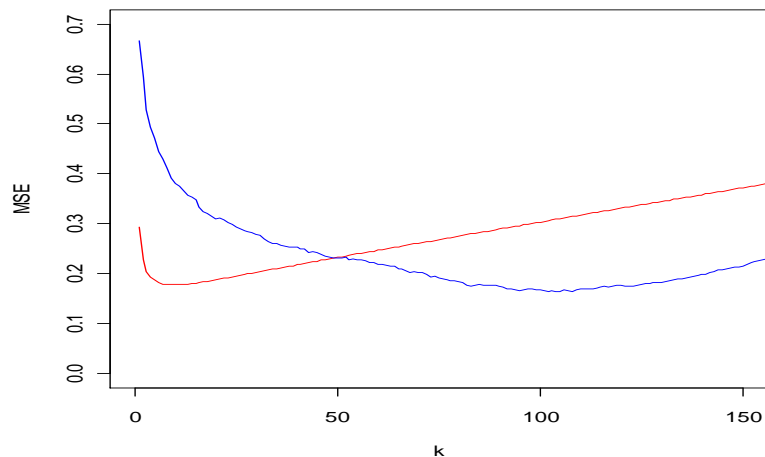


Figura 9. MSE para distribución Weibull. Método Zipf (línea roja) y EGPD (línea azul).

### 3.4. La distribución loggamma

La distribución loggamma también es comúnmente usada para modelar datos de pérdidas o severidades en riesgo operacional y seguros. La distribución loggamma  $F_{\alpha, \beta}$  tiene la siguiente función de densidad  $f_{\alpha, \beta}$ :

$$f_{\alpha, \beta} \propto (\log x)^{\beta-1} x^{-\alpha+1}$$

Para  $x > 1$ , donde  $\alpha > 0$  y  $\beta > 0$ .  $X$  se distribuye loggamma para  $X > 1$  sí y solo sí  $Y = \log X$  se distribuye gamma para  $Y > 0$ . Para valores grandes de  $x$ , se puede mostrar:

$$1 - F_{\alpha, \beta} \sim cx^{-\alpha} (\log x)^{\beta-1}.$$

Esta forma parametrizada de la distribución loggamma pertenece a MDA ( $H_1 / \alpha$ ), así que el parámetro  $\alpha$  es el que gobierna el peso de la cola. Sin embargo  $\beta$  también juega un papel importante en este modelo; si  $\beta = 1$  simplemente obtenemos la distribución Pareto. Para otros valores de  $\beta$ , lo que obtenemos es una distribución Pareto contaminada por una función de variación lenta  $(\log x)^{\beta-1}$ ; entre mayor es el valor de  $\beta$ , mayor será la contaminación.

Para ambos casos de la distribución loggamma ( $\beta = 2$  y  $10$ ), el mejor método es Zipf para valores relativamente pequeños de  $k$ , sin embargo el desempeño del estimador por el método EGPD es muy similar al del estimador de Zipf. Cuando se incrementa el valor de  $\beta$  en general, los métodos presentan mayores valores de RRMSE, lo que muestra un peor desempeño de los métodos a mayor contaminación por la función de variación lenta de la distribución loggamma. Nuevamente el estimador average Hill tiene los valores más altos de RRMSE, independiente del valor de  $\beta$  y por tal razón no se incluye en la Figura 10.

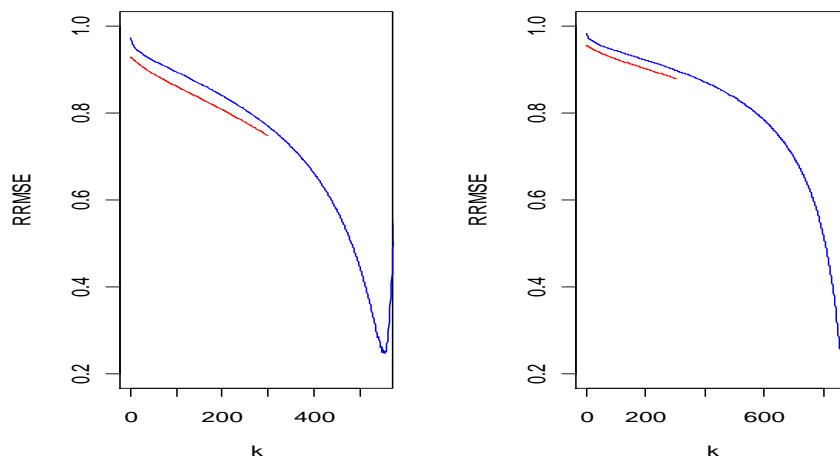


Figura 10. RRMSE para distribución loggamma con  $\beta = 2$  (izq) y  $\beta = 10$  (der). Método Zipf (línea roja) y EGPD (línea azul).

LG,  $\alpha = 1, \beta = 2$  ( $\xi = 1, \rho = 0$ )

Método	RRMSE mín.
EGPD	0.248
Zipf	0.747
Average Hill	0.956

LG,  $\alpha = 1, \beta = 10$  ( $\xi = 1, \rho = 0$ )

Método	RRMSE mín.
EGPD	0.242
Zipf	0.879
Average Hill	1.286

### 3.5 La distribución Cauchy estándar

La distribución Cauchy estándar es una distribución  $t$  – Student con un grado de libertad. El estimador average Hill es preferido para valores pequeños de  $k$  ( $k \leq 151$ ). La técnica EGPD se desempeña mejor para valores relativamente altos de  $k$  y es el método que menor RRMSE presenta.

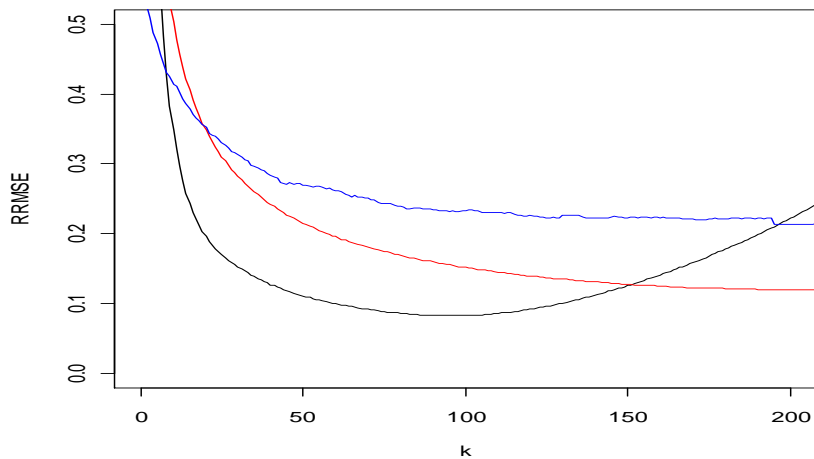


Figura 11. RRMSE para Cauchy estándar. Método average Hill (línea negra), Zipf (línea roja) y EGPD (línea azul).

Cauchy estándar ( $\xi = 1, \rho = -2$ )

Método	RRMSE mín.
EGPD	0.041
Average Hill	0.083
Zipf	0.120

### 3.6 La distribución Pareto

La distribución Pareto es la más comúnmente usada para simular distribuciones de pérdidas que presentan colas espesas. La función de distribución está dada por:

$$F_{\alpha, a}(x) = 1 - (a/x)^\alpha$$

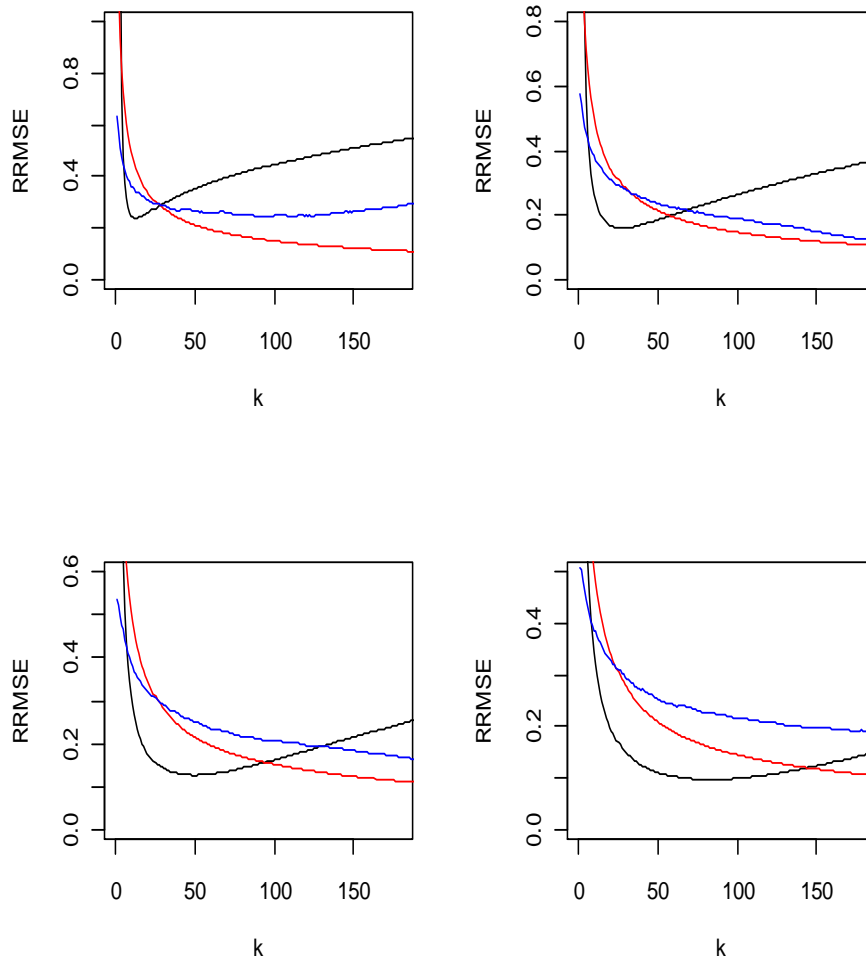
Para valores positivos de los parámetros  $a$  y  $\alpha$ ; y  $x \geq a$ . La distribución de Pareto pertenece a MDA ( $H_{1/\alpha}$ ). Como se notó en la Sección 1.2.1,  $E[X^j] = \infty$  para  $\xi \geq 1/j$ . Por tal razón hemos probado los métodos de estimación de IVE para varios valores de parámetro  $\alpha$  ( $\xi = 1/\alpha$ ), de la distribución Pareto. Para valores de  $\xi \geq 0.5$ , no es posible calcular momentos de segundo orden ni superiores. Para  $\xi \geq 1$  (1.0, 1.1 y 1.2 en este estudio), no es posible calcular ningún momento (incluso la media no es finita). En este último caso no sería posible calcular el Expected Shortfall para las distribuciones de pérdidas con estas características, denominado como los *modelos de media-infinita* (ver Nešlehová et al. (2006), y las referencias allí contenidas para una discusión más detallada de los problemas y consecuencias de estimación de altos cuantiles con este tipo de modelos). Para  $\xi \geq 1/3$ , no es posible calcular tercer momento ni superiores, pero

si la media y segundo momento, para ello tenemos un caso donde  $\zeta = 0.3$ , en el cual sí se puede calcular tercer momento, pero no el cuarto.

El estimador average Hill es el mejor método para valores relativamente pequeños de  $k$  para los diferentes valores de  $\zeta$  estudiados. Sin embargo, el estimador average Hill no es el método con menor RRMSE. Para valores pequeños de  $\zeta$  el rango del número de estadísticos ordenados superiores disminuye con respecto a la preferencia del método average Hill comparado con los otros dos estimadores, como se muestra a continuación:

$\zeta =$	0.3	0.5	0.7	1.0	1.1	1.2
$k \leq$	27	57	94	144	168	180

El estimador de Zipf también se desempeña bien para todos los casos. Cuando  $\zeta = 0.3$ , 0.5 y 1.1 el método de Zipf tiene el menor RRMSE (0.085, 0.082 y 0.085 respectivamente). El método EGPD tiene el menor RRMSE cuando  $\zeta = 0.7, 1.0$  y 1.2 (0.052, 0.036 y 0.030 respectivamente).



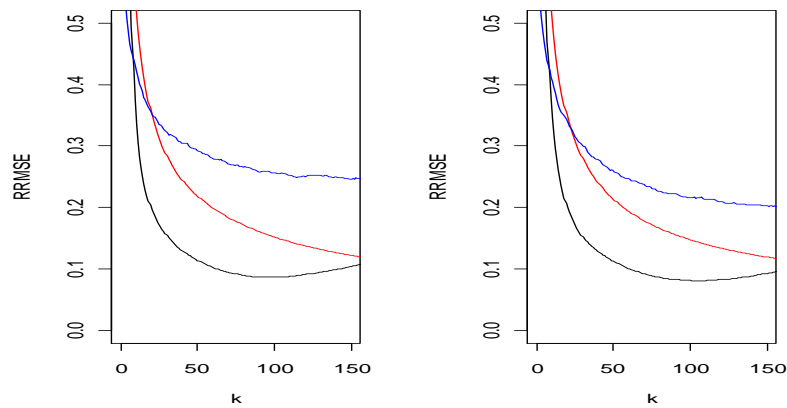


Figura 12. RRMSE para la distribución Pareto con  $\zeta = 0.3$  (arriba izq),  $\zeta = 0.5$  (arriba der),  $\zeta = 0.7$  (centro izq),  $\zeta = 1.0$  (centro der),  $\zeta = 1.1$  (abajo izq),  $\zeta = 1.2$  (abajo der). Método average Hill (línea negra), Zipf (línea roja) y EGPD (línea azul).

Pareto, ( $\zeta = 0.3, \rho = -\infty$ )

Método	RRMSE mín.
Zipf	0.085
Average Hill	0.237
EGPD	0.243

Pareto, ( $\zeta = 0.5, \rho = -\infty$ )

Método	RRMSE mín.
Zipf	0.082
EGPD	0.087
Average Hill	0.160

Pareto, ( $\zeta = 0.7, \rho = -\infty$ )

Método	RRMSE mín.
EGPD	0.052
Zipf	0.086
Average Hill	0.126

Pareto, ( $\zeta = 1.0, \rho = -\infty$ )

Método	RRMSE mín.
EGPD	0.036
Zipf	0.081
Average Hill	0.097

Pareto, ( $\zeta = 1.1, \rho = -\infty$ )

Método	RRMSE mín.
Zipf	0.085
Average Hill	0.087
EGPD	0.140

Pareto, ( $\zeta = 1.2, \rho = -\infty$ )

Método	RRMSE mín.
EGPD	0.030
Average Hill	0.081
Zipf	0.084

### 3.7 Comentarios

Si el criterio de selección de un estimador fuese el mínimo valor de RRMSE (o MSE dependiendo del caso), para estimar el IVE, en general, podríamos recomendar el método EGPD en conjunto con el estimador average Hill si los datos se distribuyen  $t$  – Student (casos donde  $\zeta < 1$ ). Para el caso Pareto se recomienda usar el estimador de Zipf en conjunto con el método EGPD. El método EGPD tiene el mínimo RRMSE para datos que se distribuyen Cauchy estándar y loggamma (donde  $\zeta = 1$ ). Los métodos anteriormente revisados están diseñados para desempeñarse bien en casos cuando  $\zeta > 0$ .

Sin embargo, para los casos en que los datos se distribuyen lognormal y Weibull (donde  $\xi = 0$ ), el método EGD<sub>P</sub> presenta el menor RRMSE; aunque los gráficos de MSE vs.  $k$  lucen un poco extraño para medir el desempeño de estos estimadores en los casos de loggamma y lognormal (donde  $\rho = 0$ ), por tal razón es recomendable también analizar métodos para estimación de  $\xi \in \mathbb{R}$  y podría ser objeto de un estudio más adelante.

Recordemos que la distribución subyacente de los datos es desconocida en la práctica, pero de acuerdo a la simulación, el método EGD<sub>P</sub> es la técnica que mejor se desempeña cuando  $-1 < \rho \leq 0$ . En otras palabras, se corrobora la teoría de que este método es una buena técnica de reducción de sesgo presentado en el gráfico de Hill, y podría ser usado en el área de seguros y riesgo operacional cuando se asume que las severidades siguen una distribución loggamma, lognormal o Weibull con fines de simulación y estimación de cuantiles altos (como VaR al 99% ó 99.9%). Otra idea para saber qué método utilizar en datos de pérdida históricos, es realizar una prueba de bondad de ajuste y observar que distribución paramétrica es la que más se aproxima a los datos empíricos y así seleccionar el estimador que mejor se desempeña de acuerdo a las observaciones realizadas en este documento y están basados en los resultados obtenidos de la simulación.

## Caso: Danish Fire Data

Aplicamos los métodos revisados a un caso particular. Los datos corresponden a 2167 reclamos en seguro contra el fuego en Dinamarca (Danish Fire Data), donde las pérdidas se expresan en millones de Coronas Danesas (DKM) desde marzo 1 de 1980 hasta diciembre 12 de 1990. Los datos están disponibles en [www.ma.hw.ac.uk/~mcneil/](http://www.ma.hw.ac.uk/~mcneil/) y ha sido previamente estudiado por varios autores bajo el marco de EVT. Empezaremos con un análisis exploratorio de los datos. La Figura 13 muestra la serie de tiempo para el Danish Fire data.

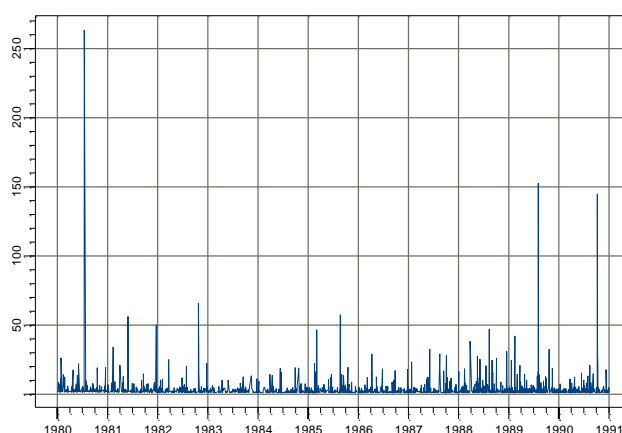


Figura 13. Danish Fire Data

A continuación se muestra el resumen del análisis y se observa que la distribución de las pérdidas es sesgada a la derecha y presenta alta kurtosis (buen ejemplo para estudio de colas espesas).

Danish data	Pérdida en DKM
Min	1.00
Primer cuartil	1.32
Mediana	1.77
Media	3.38
Sesgo	18.76
Kurtosis	483.76
Tercer cuartil	2.96
Max	263.25
$\hat{Q}_{0.99}$	26.04

El cuantil al 99% de los datos empíricos es igual a 26.04 DKM. Más adelante calcularemos cuantiles a este nivel con los diferentes métodos estudiados para compararlo con este dato.



La Figura 14 muestra el Pareto quantile plot:

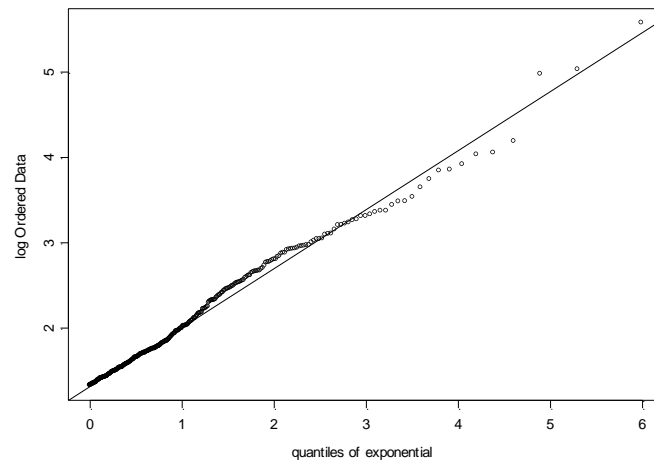


Figura 14. Pareto quantile plot para Danish data

El Pareto quantile plot nos dice que los datos exhiben colas espesas. Se tendería a pensar que la distribución subyacente es muy cercana a una distribución Pareto debido a que la curva graficada en el Pareto quantile plot es muy cercano a una recta en su parte derecha. La pendiente de la línea justada en el Pareto quantile plot es una buena aproximación del IVE cuando se asume que los datos siguen una distribución Pareto (como lo habíamos mencionado anteriormente). Para este caso la pendiente es igual 0.722 como primera aproximación del IVE para Danish Fire Data. De todas maneras no estamos completamente seguros de que la distribución subyacente de los datos siga una distribución Pareto. A continuación se observa el gráfico de función de autocorrelación (ACF, por sus siglas en inglés).

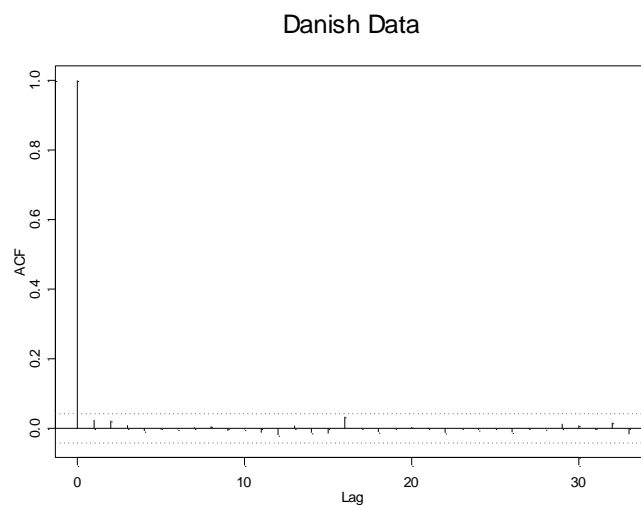


Figura 15. ACF de Danish data

Como se observa en la Figura anterior los datos no exhiben correlación serial. Esto sucede a menudo con datos del sector seguros y se puede presumir que son observaciones iid, sin embargo esto no es muy común con los datos financieros. Vea Resnick (1997) para pruebas adicionales de independencia aplicados a Danish data. La siguiente Figura muestra los gráficos de los métodos aplicados al Danish Fire Data para estimar  $\zeta$ .

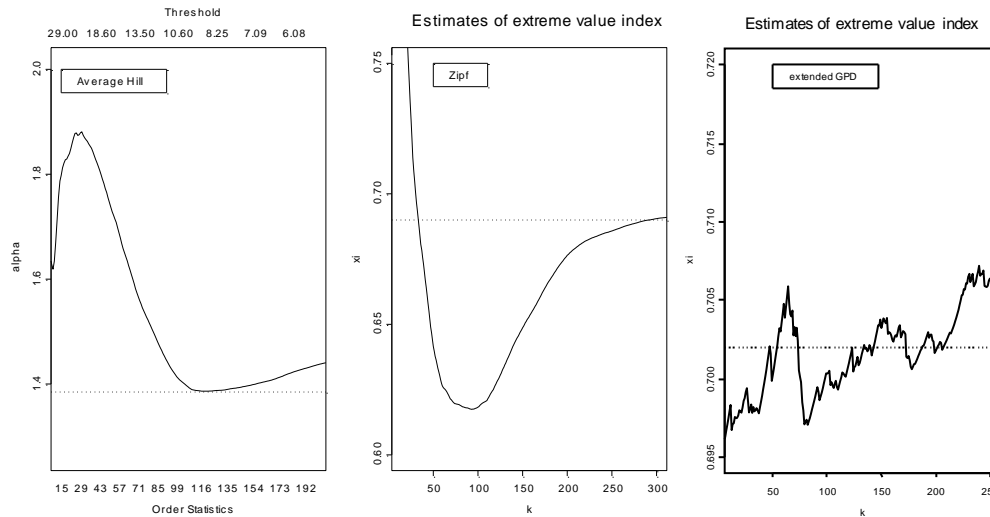


Figura 16. Los métodos aplicados al caso Danish data.

La estabilidad en los gráficos para los métodos average Hill, Zipf y EGPD dan los siguientes resultados:

$$\hat{\alpha}^{\text{vH}} = 1.385, \hat{\zeta}_{120,2167}^{\text{vH}} = 0.722 (k = 120),$$

$$\hat{\zeta}_{290,2167}^{\text{Zipf}} = 0.690, (k = 290),$$

$$\hat{\zeta}_{290,2167}^{\text{EGPD}} = 0.702 (k = 72).$$

También estamos interesados en estimar un cuantil alto  $Q(1 - p)$  para Danish data, donde  $p$  es un número muy pequeño entre 0 y 1. Para ello usaremos el estimador de Weissman para la mayoría de los casos (excepto para el método EGPD).

*Estimador de Weissman.* Asuma que la linealidad en el Pareto quantile plot persiste desde las  $k$  observaciones más grandes (hasta infinito), entonces podemos usar la línea con ecuación:

$$y = \log X_{n-k,n} + \hat{\xi}_{k,n}^H \left( x + \log \left( \frac{k+1}{n+1} \right) \right),$$

anclado en el punto  $\left( -\log \left( \frac{k+1}{n+1} \right), \log X_{n-k,n} \right)$ . Sea  $x = -\log p$  para obtener el estimado de  $Q(1-p)$ :

$$\begin{aligned} \hat{Q}(1-p) &= \exp \left( \log X_{n-k,n} + \hat{\xi}_{k,n}^H \log \left( \frac{k+1}{(k+1)p} \right) \right) \\ &= X_{n-k,n} \left( \frac{k+1}{(k+1)p} \right)^{\hat{\xi}_{k,n}^H}, \quad k=1, \dots, n-1. \end{aligned}$$

*Estimador de cuantil para el método EGPD.* La estimación se realiza mediante el ajuste de los excesos relativos  $Y_j$ ,  $j = 1, \dots, k$ , (por encima de un umbral  $X_{n-k}$ ) a una distribución Pareto, resultando en el siguiente estimado de cola para  $P(X > x)$  y para valores grandes de  $x$ :

$$\hat{p}_{k,x} = \frac{k+1}{n+1} \bar{G}_P \left( X_{n-k}, \hat{\xi}, \hat{\delta}, \hat{\rho} \right)$$

Donde  $\bar{G}_P$  denota la función de supervivencia de la distribución Pareto. Fijando  $\hat{p}_{k,x}$  en un valor muy pequeño (por ejemplo 0.01 para estimar  $\hat{Q}(0.99)$ ), se puede resolver numéricamente para  $x$ , y así obtener un estimador de cuantil extremo.

*Estimador de cuantil para el método POT.* Bajo los supuestos mencionados en la Sección 1.2.1 de este documento, tenemos para  $x \geq u$ :

$$\begin{aligned} \bar{F}(x) &= P(X > u) \bar{P}(X > x | X > u) \\ &= \bar{F}(u) \bar{P}(X - u > x - u | X > u) \\ &= \bar{F}(u) \bar{F}(x - u) \\ &= \bar{F}(u) \left( 1 + \xi \frac{x - u}{\beta} \right)^{-1/\xi}. \end{aligned}$$

La formula se puede invertir para obtener un cuantil alto de la distribución subyacente:

$$Q(1-p) = u + \frac{\beta}{\xi} \left( \left( \frac{p}{\bar{F}(u)} \right)^{-\xi} - 1 \right)$$

Para ello requerimos un estimado de  $\bar{F}(\cdot)$ , entonces se propone usar el estimador  $Nu/n$  (ver McNeil et al. (2005), y las referencias allí contenidas).  $\zeta$  y  $\beta$  son los parámetros de forma y escala de GPD. Para estimar los cuantiles por el método POT usamos la librería QRMLib (*Quantitative Risk Management library*) para S-PLUS que se puede encontrar en [www.ma.hw.ac.uk/~mcneil/book/QRMLib.html](http://www.ma.hw.ac.uk/~mcneil/book/QRMLib.html) con un previo registro.

Usamos la siguiente información (que corresponde a  $X_{n-k,n}$ ) para los métodos average Hill, Zipf y EGPD respectivamente para estimar  $Q(0.99)$  para Danish Fire Data:

- $X_{2047,2167} = 8.72$ ,
- $X_{1887,2167} = 4.61$ ,
- $X_{2095,2167} = 13.35$ ,  $\hat{\delta} = -0.15$ ,  $\hat{\rho} = -4.6$ .

De esta manera calculamos el cuantil para cada uno de los métodos:

- $\hat{Q}^{(vH)}(0.99) = 30.19$ ,
- $\hat{Q}^{(H)}(0.99) = 15.09$ ,
- $\hat{Q}^{(EGPD)}(0.99) = 24.90$ .

#### 4.1 Comentarios

Los métodos analizados en este documento arrojan un IVE entre 0.690 y 0.722. Mediante el método POT, McNeil (1997) obtuvo un estimado de  $\zeta$  igual a 0.684 (con error estándar de 0.27) para un umbral de 20 DKM y  $k = 36$ . Csörgő y Viharos (1998) también estimaron el IVE por varios métodos para Danish Data y obtuvieron los siguientes resultados: 0.716, 0.713, 0.719, 0.717 y 0.719 (con  $M = 0.000323$ ).

Los resultados obtenidos en este documento para la estimación del IVE son similares a los obtenidos por Csörgő y Viharos y también a los de McNeil cuando usa un umbral de 20 DKM. Sin embargo, cuando el umbral es 10 DKM, McNeil obtiene un estimado de  $\zeta$  igual a 0.497 (con error estándar de 0.14).

Usando la información del artículo de McNeil (1997), estimamos  $Q(0.99)$  usando el método POT. Para un umbral de 20 DKM es 25.8 DKM; y 27.3 DKM para un umbral de 10 DKM. Recordemos que al principio de esta sección se estimó el cuantil empírico de 26.04 DKM y el método analizado en este documento que más se aproxima es el EGPD con un valor de 24.9 DKM. Recientemente Degen et al. (2007) utilizan la aproximación de las pérdidas en exceso a una distribución g-and-h (en lugar de una distribución GPD como lo hace el método POT) y estiman  $Q(0.99)$  para Danish Fire Data en 27.32 DKM. Para más detalles de esta novedosa aproximación se remite al lector al artículo de Degen et al. y las referencias allí contenidas.

## Conclusiones y futura investigación

El índice de valor extremo es el parámetro que controla la forma de la cola de la distribución. Los administradores de riesgo no solamente están interesados en estimar capital regulatorio sino también capital económico. Las pérdidas esperadas deberían ser cubiertas con capital regulatorio, mientras que el capital económico cubriría pérdidas inesperadas por eventos extremos. Un estimado confiable del parámetro del índice de valor extremo conlleva a estimados confiables de altos cuantiles (como medida de riesgo) y por ende estimados razonables de cargos de capital.

Los administradores de riesgo deben tener en cuenta que los eventos extremos pueden ocurrir y causar grandes pérdidas. El problema es que los valores extremos escondidos en una cola de una distribución no son fáciles de detectar. Una herramienta para solucionar este problema es EVT. Sin embargo, si EVT es la herramienta elegida para tratar los valores extremos del negocio y estimar medidas de riesgo, el área de administración de riesgos debe considerar el problema de la estimación del “parámetro de forma”. Es bien sabido que el VaR no es una medida coherente al riesgo y puede conllevar subestimaciones de valores de riesgo de mercado y cuando se usa una medida coherente al riesgo (como el Expected Shortfall o tail VaR) se obtienen cargos de capital mayores (y a veces no razonables).

Lo que sugerimos a los reguladores y profesionales en riesgo es comenzar a probar modelos EVT para cuantificar riesgos financieros (mercado, crédito y operacional) y en el sector de seguros. Pero se debe mitigar las trampas al usar EVT. Una de esas trampas es la selección del umbral al aplicar el método POT para estimar el parámetro de forma, un problema aún no resuelto. Sin embargo, en este documento probamos algunos métodos para estimar el IVE (o el parámetro de forma).

La técnica más comúnmente usada para estimar el IVE es el estimador de Hill. Aunque el estimador de Hill es un estimador consistente para  $\zeta$  para datos iid (bajo ciertas condiciones también para datos no iid), este método tiene algunos problemas:

- (1) Escoger el valor de  $k$  del gráfico de Hill no es una tarea fácil.
- (2) El gráfico de Hill puede mostrar alguna volatilidad y si la distribución subyacente de los datos es muy diferente a una Pareto, el gráfico puede exhibir sesgo.
- (3) El estimador de Hill no es invariante en localización.

En el punto (2) revisamos el estimador average Hill como una posible solución para la alta volatilidad. Resnick y Stărică (1997) propone el uso de este método como una base para técnicas de bootstrap para corregir el problema de sesgo. El estimador de Zipf puede ser visto también como una técnica de suavización del gráfico de Hill para estimar el IVE. Sin embargo, su varianza asintótica es el doble de la varianza del estimador de Hill. Pero una ventaja de este método es que los residuos del Pareto quantile plot contiene información el cual podría ser usado para mitigar el sesgo en los estimados cuando la cola de la distribución no es Pareto. Los estimadores de Hill y de

Zipf pertenecen a una clase más grande de kernel el cual podría ser interesante para posteriores estudios. Respecto al sesgo, probamos el método EGPD el cual reduce sesgo comparado con el estimador de Hill.

Desafortunadamente no podemos responder a la pregunta de “cuál es el mejor método a usar si existe para calcular el IVE?” la selección *optima* del umbral (o selección de  $k$ ) no está resuelto aún, como lo mencionamos anteriormente. Esta selección optima puede ser obtenida solo bajo algunas propiedades precisas de segundo orden en la función de variación lenta  $L$ , la cual no podemos inferir de los datos. Pero podemos dar una guía cuando trabajamos con datos si conocemos sus funciones de distribución. Por ejemplo, en el sector de seguros de no vida es común usar las distribuciones loggamma y lognormal para modelar las distribuciones de pérdidas. Sugerimos usar la técnica de EGPD (técnica de reducción de sesgo) en estos casos para estimar el IVE. En general se recomienda usar este método en conjunto del estimador de average Hill o de Zipf. A partir de los resultados de la simulación pudimos comprobar el buen desempeño de la técnica EGPD, aunque existe otra familia de estimadores por probar.

Existe también la posibilidad de analizar los métodos para estimar el IVE cuando  $\xi \in \mathbb{R}$ . De igual manera existen métodos robustos para estimación del IVE que se podrían probar en simulaciones como las mostradas en este documento. Finalmente pensamos que un método novedoso en EVT es la aproximación de las pérdidas en exceso a una distribución g-and-h más que a una GPD para calcular valores altos de cuantiles como medida de riesgo.

## Referencias

- Beirlant, J., Dierckx G., Goegebeur, Y. y Matthys G. (1999) Tail index estimation and an exponential regression model. *Extremes* **2**, 177-200.
- Beirlant, J. Goegebeur, Y., Segers, J. and Teugels, J. (2004) *Statistics of Extremes: Theory and Applications*. John Wiley Sons.
- Beirlant, J., Joossens E. and Segers J. (2005) Unbiased tail estimation by an extensión of the generalized Pareto distribution. *CentER Discussion Paper* 2005-112.
- Bingham, N., Goldie, C. y Teugels, J. (1987) *Regular Variation*. Cambridge University Press.
- Csörgő, S. y Viharos L. (1998) Estimating the tail index. In Szyszkowicz, B. (Ed.), *Asymptotic Methods in Probability and Statistics*, 833-881. Amsterdam: North-Holland.
- Degen, M., Embrechts, P. y Lambrigger, D. (2007) The quantitative modeling of operational risk: between g-and-h and EVT. *ASTIN Bulletin* **37**: 265-291.
- Drees, H., de Haan, L. y Resnick, S. I. (2000) How to make a Hill plot. *Annals of Statistics* **28**, 254-274.
- Embrechts, P., Klüppelberg, C. y Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin.
- Gomes, M. I. y Martins M. J. (2002) "Asymptotically unbiased" estimators of the tail index. *Extremes* **5**, 1-31.
- Kratz M. y Resnick S. I. (1996) The qq-estimator and heavy tails. *Stochastic Models*, **12**: 699-724.
- McNeil, A. J. (1997) Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin* **27**: 117-137.
- McNeil, A. J., Frey, R. y Embrechts, P. (2005) *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, Princeton.
- McNeil, A., y T. Saladin (1997) "The Peaks over Thresholds Method for Estimating High Quantiles of Loss Distributions", in Proceedings of XXVIIth International *ASTIN Colloquium*, pp. 23-43, Cairns, Australia.
- Matthys, G. y Beirlant, J. (2000) Adaptive threshold selection in tail index estimation. In: Embrechts P. (Ed.), *Extremes and Integrated Risk Management*, pp. 37-48, Risk Books, London.
- Nešlehová, J., Chavez-Demoulin, V., Embrechts, P., (2006) Infinite-mean models and the LDA for operational risk. *Journal of Operational Risk*, pp. 3-25.

Reiss, R.-D. y Thomas, M. (1997) *Statistical Analysis of Extreme Values*. Basel: Birkhäuser.

Resnick, S. I. (1987) *Extreme Values, Regular Variation and Point Processes*. Springer, New York.

Resnick, S. I. (1997) Discussion of the Danish data on large fire insurance losses. *Astin Bulletin* 27: 139-151.

Resnick, S. I. y Stărică, C. (1997) Smoothing the Hill estimator. *Advances in Applied Probability* **29**: 271-293.

Schultze, J. y Steinebach, J. (1996) On least squares estimates of an exponential tail coefficient. *Statist. Decisions*, **14**, 353-372.